AE 17: Hypothesis tests for independence

Your name

2021-10-19

```
library(tidyverse)
library(tidymodels)
```

```
yawn <- read_csv("data/yawn.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   result = col_character(),
##   group = col_character()
## )
```

Bulletin

- Due today:
  - lab 06
- Due Thursday:
  - prep quiz 06
- Upcoming
  - proposals back to you by end of Wednesday

Learning goals

Use and understand simulation-based methods to ...

- test a claim about a population proportion
- test a claim about independence between two groups

Test for independence

- First let's, watch the experiment from *Mythbusters*.
- Let `$t$` be the treatment group who saw a person yawn, `$c$` be the control group who did not see anyone yawn, and `$p$` be the proportion of people who yawned.

Exercise 1[1]

We want to use simulation-based inference to assess whether or not yawning and seeing someone yawn are independent.

[1] Simulation activity from Data science in a box

- State the null and alternative hypotheses in words:
- Select the appropriate null and alternative hypotheses written in mathematical notation:

a. `$H_0: p_t = p_c \text{ vs. }H_a: p_t < p_c$`

b. `$H_0: p_t = p_c \text{ vs. }H_a: p_t > p_c$`

c. `$H_0: p_t = p_c \text{ vs. }H_a: p_t \neq p_c$`

d. `$H_0: \hat{p}_t = \hat{p}_c \text{ vs. }H_a: \hat{p}_t < \hat{p}_c$`

e. `$H_0: \hat{p}_t = \hat{p}_c \text{ vs. }H_a: \hat{p}_t > \hat{p}_c$`

f. `$H_0: \hat{p}_t = \hat{p}_c \text{ vs. }H_a: \hat{p}_t \neq \hat{p}_c$`

Exercise 1.5

Type I and Type II error

Truth	Reject null	Fail to reject
Null is true	Type 1 error	Good decision
Null is false	Good decision	Type 2 error

What does type I error look like in this case?

What does type II error look like?

Is type I or type II error more worrisome here?

How might we mitigate the more dangerous error?

Exercise 2

Before using R to construct the null distribution, let's generate the null distribution using playing cards! See AE-17 for the simulation instructions. You can also find them in the `README` of this application exercise.

Exercise 3

Uncomment the code to see read in the data from the class and visualize the null distribution.

```
#sim_data <- read_csv("https://sta199-f21-001.netlify.app//appex/data/yawn-sim.csv")
```

```
#ggplot(data = sim_data, mapping = aes(x = diff_in_prop)) +
#  geom_histogram(binwidth = 0.05) +
#  labs(title = "Your Results: Difference in Proportion of Yawners")
```

- What is the approximate center of the distribution? Is this what you expected? Why or why not?
- The observed difference in proportions from the Mythbusters episode is 0.0441. Based on your simulated distribution, do yawning and seeing someone yawn appear to be dependent?

Exercise 4

Let's use the data from the *Mythbusters* episode and simulation-based inference in R to test this claim. Based on their experiment, do yawning and seeing someone yawn appear to be dependent?

Evaluate this question using a simulation based approach. We will use the same null and alternative hypotheses as before. The data from *Mythbusters* is available in the `yawn` data frame.

- Fill in the code below to generate the null distribution. Uncomment the code once it is complete.

```
set.seed(101821)
#null_dist <- yawn %>%
#  specify(response = ____, explanatory = _____, success = "yawn") %>%
#  hypothesize(null = "______") %>%
#  generate(100, type = "permute") %>%
#  calculate(stat = "______",
#            order = c("trmt", "ctrl"))
```

- Visualize the null distribution and shade in the area used to calculate the p-value.

```
# add code
```

- Calculate p-value. Then use the p-value to make your conclusion using a significance level of 0.1.

```
# add code
```

Exercise 5

Do you believe the conclusions from this experiment? Why or why not?